



King's Research Portal

DOI:

[10.1016/j.molonc.2015.03.002](https://doi.org/10.1016/j.molonc.2015.03.002)

Document Version

Publisher's PDF, also known as Version of record

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Schouten, P. C., Grigoriadis, A., Kuilman, T., Mirza, H., Watkins, J. A., Cooke, S. A., van Dyk, E., Severson, T. M., Rueda, O. M., Hoogstraat, M., Verhagen, C. V. M., Natrajan, R., Chin, S-F., Lips, E. H., Kruizinga, J., Velds, A., Nieuwland, M., Kerkhoven, R. M., Krijgsman, O., ... Linn, S. C. (2015). Robust BRCA1-like classification of copy number profiles of samples repeated across different datasets and platforms. *Molecular Oncology*, 9(7), 1274-86. <https://doi.org/10.1016/j.molonc.2015.03.002>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

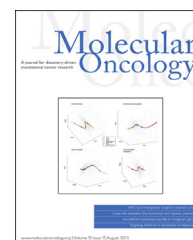
- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/molonc

CrossMark

Robust *BRCA1*-like classification of copy number profiles of samples repeated across different datasets and platforms

Philip C. Schouten^a, Anita Grigoriadis^b, Thomas Kuilman^c, Hasan Mirza^b, Johnathan A. Watkins^b, Saskia A. Cooke^b, Ewald van Dyk^d, Tessa M. Severson^a, Oscar M. Rueda^e, Marlous Hoogstraat^{a,d,f,g}, Caroline V.M. Verhagen^h, Rachael Natrajanⁱ, Suet-Feung Chin^e, Esther H. Lips^a, Janneke Kruizinga^j, Arno Velds^j, Marja Nieuwland^j, Ron M. Kerkhoven^j, Oscar Krijgsman^c, Conchita Vens^h, Daniel Peeper^c, Petra M. Nederlof^k, Carlos Caldas^{e,l,m}, Andrew N. Tutt^b, Lodewyk F. Wessels^{d,n}, Sabine C. Linn^{a,o,p,*}

^aDepartment of Molecular Pathology, Netherlands Cancer Institute, Amsterdam, The Netherlands

^bBreakthrough Breast Cancer Research Unit, Department of Research Oncology, Guy's Hospital, King's College London School of Medicine, London, United Kingdom

^cDivision of Molecular Oncology, Netherlands Cancer Institute, Amsterdam, The Netherlands

^dDepartment of Molecular Carcinogenesis, Netherlands Cancer Institute, Amsterdam, The Netherlands

^eCancer Research UK Cambridge Research Institute, Li Ka Shing Centre, Cambridge, UK

^fDepartment of Medical Oncology, University Medical Center Utrecht, Utrecht, The Netherlands

^gNetherlands Center for Personalized Cancer Treatment, Utrecht, The Netherlands

^hDivision of Biological Stress Response, Netherlands Cancer Institute, Amsterdam, The Netherlands

ⁱThe Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, London, UK

^jGenomics Core Facility, Netherlands Cancer Institute, Amsterdam, The Netherlands

^kDepartment of Pathology, Netherlands Cancer Institute, Amsterdam, The Netherlands

^lDepartment of Oncology, University of Cambridge, Addenbrooke's Hospital, Cambridge, UK

^mCambridge Experimental Cancer Medicine Centre and NIHR Cambridge Biomedical, Research Centre, Cambridge University Hospitals NHS, Cambridge, UK

ⁿFaculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands

^oDepartment of Pathology, University Medical Center Utrecht, Utrecht, The Netherlands

^pDivision of Medical Oncology, Netherlands Cancer Institute, Amsterdam, The Netherlands

Abbreviations: aCGH, array Comparative Genomic Hybridization; BAC, Bacterial Artificial Chromosome; BAC32K, Bacterial Artificial Chromosome aCGH, 32K platform; BAC3K, Bacterial Artificial Chromosome aCGH, 3.2K platform; BRCA1, Breast Cancer Early Onset 1; CN, Copy number; DNA, Deoxyribonucleic acid; dsDNA, double-stranded DNA; FFPE, Formalin Fixed Paraffin Embedded; hg 18, human reference genome version 18; hg19, human reference genome version 19; MIP, Molecular Inversion Probe; NG135, Nimblegen 135k oligonucleotide aCGH; NG720, Nimblegen 720K oligonucleotide aCGH; NGS, Low coverage next generation sequencing; SNP6, Affymetrix SNP6 array; SNR, Signal to Noise Ratio; VN, Variance of the Noise.

* Corresponding author. Departments of Medical Oncology and Molecular Pathology at Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands. Tel.: +31 20 512 2951.

E-mail address: s.linn@nki.nl (S.C. Linn).

<http://dx.doi.org/10.1016/j.molonc.2015.03.002>

1574-7891/© 2015 The Authors. Published by Elsevier B.V. on behalf of Federation of European Biochemical Societies. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

ARTICLE INFO

Article history:

Received 26 November 2014

Received in revised form

1 March 2015

Accepted 11 March 2015

Available online 20 March 2015

Keywords:

BRCA1

Breast cancer

Classification

Copy number aberration profiles

ABSTRACT

Breast cancers with BRCA1 germline mutation have a characteristic DNA copy number (CN) pattern. We developed a test that assigns CN profiles to be 'BRCA1-like' or 'non-BRCA1-like', which refers to resembling a BRCA1-mutated tumor or resembling a tumor without a BRCA1 mutation, respectively. Approximately one third of the BRCA1-like breast cancers have a BRCA1 mutation, one third has hypermethylation of the BRCA1 promoter and one third has an unknown reason for being BRCA1-like. This classification is indicative of patients' response to high dose alkylating and platinum containing chemotherapy regimens, which targets the inability of BRCA1 deficient cells to repair DNA double strand breaks. We investigated whether this classification can be reliably obtained with next generation sequencing and copy number platforms other than the bacterial artificial chromosome (BAC) array Comparative Genomic Hybridization (aCGH) on which it was originally developed. We investigated samples from 230 breast cancer patients for which a CN profile had been generated on two to five platforms, comprising low coverage CN sequencing, CN extraction from targeted sequencing panels (CopywriteR), Affymetrix SNP6.0, 135K/720K oligonucleotide aCGH, Affymetrix Oncoscan FFPE (MIP) technology, 3K BAC and 32K BAC aCGH. Pair-wise comparison of genomic position-mapped profiles from the original aCGH platform and other platforms revealed concordance. For most cases, biological differences between samples exceeded the differences between platforms within one sample. We observed the same classification across different platforms in over 80% of the patients and kappa values of at least 0.36. Differential classification could be attributed to CN profiles that were not strongly associated to one class. In conclusion, we have shown that the genomic regions that define our BRCA1-like classifier are robustly measured by different CN profiling technologies, providing the possibility to retro- and prospectively investigate BRCA1-like classification across a wide range of CN platforms.

© 2015 The Authors. Published by Elsevier B.V. on behalf of Federation of European Biochemical Societies. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Breast cancer arising in patients with a germline BRCA1 mutation are thought to be genomically unstable due to the impairment of error-free homologous recombination DNA repair in which BRCA1 has a role (Venkitaraman, 2009; Vollebergh et al., 2012). DNA copy number (CN) profiles provide a snapshot of a result of genomic instability in cancer, namely the CN aberrations. The copy number profiles of patients with a BRCA1 mutation have specific gains and losses (Alvarez et al., 2005; Tirkkonen et al., 1997; Wessels et al., 2002). We previously developed a shrunken centroids classifier which uses 371 genomic regions to assign a CN profile to the BRCA1-like (sharing characteristics of BRCA1 mutated breast cancer) or non-BRCA1-like phenotype (Vollebergh et al., 2011). This classifier not only identifies germline BRCA1-mutated cases (approximately 1/3 of the BRCA1-like tumors) but also enriches for tumors with other mechanisms of BRCA1 inactivation, for example promoter hypermethylation (approximately 1/3 of the BRCA1-like tumors, mutually exclusive with BRCA1 mutation) (Joosse et al., 2011; Vollebergh et al., 2011; Lips et al., 2011) which can confer to non-familial cases a tumor phenotype that is similar to BRCA1 mutation carriers. Alternative modes of BRCA1 inactivation and similarity of these tumors to BRCA1-mutated tumors have been observed in other datasets as well (Turner et al., 2004; Esteller et al., 2000; Alvarez et al., 2005; Tung et al., 2010; Cancer Genome Atlas Network, 2012) and has been referred to as 'BRCAness' (Turner et al., 2004). The

cases with unknown cause for being classified as BRCA1-like may thus be subject to BRCA1 dysfunction due to yet unidentified causes, or reflect a broader pathway dysfunction. Subsequently, we demonstrated that BRCA1-like patients benefit significantly more from high dose DNA double strand break-inducing chemotherapy, containing both platinum and alkylating agents, than from a conventional second generation chemotherapy regimen (Vollebergh et al., 2011). Two follow-up studies with different chemotherapy regimens demonstrated that BRCA1-like patients benefit also from tandem high dose (both including alkylating agents, one including platinum) compared to conventional, and from tandem high dose compared to dose dense chemotherapy, underlining the clinical relevance of the BRCA1-like profile (Schouten et al., 2015, 2014, 2013b). Technological advances in experimental platforms have provided many datasets to study BRCA1-like profiles next to those generated on the original BAC (BAC3K) platform and 135k oligonucleotide aCGH (NG135), on which we reported in a previous manuscript (Schouten et al., 2013a). Given this reported reproducibility between different CN profiling platforms, we investigated whether BRCA1-like classification of CN profiles of repeated samples could be reliably obtained across multiple platforms (Baumbusch et al., 2008; Curtis et al., 2009; Hester et al., 2009; Krijgsman et al., 2012; Schouten et al., 2013a; Wicker et al., 2007). For this study we compared data from samples for which data from at least two of the following platforms were available: low coverage genome-wide sequencing,

[illegible]

TruSeq® DNA LT Sample Preparation kit (Illumina). The double-stranded DNA input amount was lower than advised by the TruSeq protocol, so we used up to 250 ng of double-stranded DNA, such that 2.5 times less adapter concentration was used than prescribed in the TruSeq protocol. During enrichment PCR, 10 cycles were necessary to obtain enough yield for sequencing. All DNA libraries were analyzed on a Bio-Analyzer system (Agilent Technologies) using the DNA7500 chips for determining the molarity. Up to ten uniquely indexed samples were pooled equimolarly to give a final concentration of 10 nM. Pools were then sequenced using an Illumina HiSeq2000 machine to a coverage of 0.5×. This was done in one lane of a single-end 50 bp run according to manufacturer's instructions.

Reads were aligned to the reference genome (hg19) using the BWA backtrack algorithm (Li and Durbin, 2009) and counted in 20 kb non-overlapping bins to obtain reliable number of counts to derive copy number from within these bins. The bin counts were corrected for GC bias using a loess fit (Benjamini and Speed, 2012). The mappability value of each bin is precomputed by summarizing the alignment results of all possible 51-mers from the reference sequence. A linear model intercepting 0 was used to fit the loess-corrected count data to the mappability values (Supplementary Figure 1 shows an example loess fit to correct for mappability and GC biases). The slope of this fit, multiplied with the mappability value for each bin, provides the bin's reference value that is used to calculate the final log2 copy number ratios. ENCODE (ENCODE Project Consortium et al., 2012) blacklisted regions and bins with a mappability <0.2 were excluded from the final dataset. Subsequently the log2 copy number ratios were mapped to the original BAC clone locations, which were extended to 1 MB to capture a sufficient number of reads for every BAC clone.

2.5. Targeted sequencing and data processing

Three µg of DNA from N4+ samples was used to prepare paired-end fragment libraries using a genomic DNA library preparation kit (Illumina). The libraries were hybridized to a SureSelect custom-based bait library (Agilent) enriching for 565 genes involved in DNA repair and cancer ("DNA repair-ome"). After washing the captured DNA was amplified. Enriched libraries were barcoded, pooled and sequenced on an Illumina HiSeq 2000 machine using a 2×75 bp paired-end protocol. Reads were filtered for quality and aligned to the human genome (GRCh37/hg19) using Samtools.

Bar-coded sequence libraries for the NAC samples were generated based on (Vermaat et al., 2012). 300–600 ng of input DNA was used (Harakalova et al., 2011). These pools were enriched for 1977 ("Cancer mini-genome") cancer-related genes using SureSelect technology. Enriched libraries were sequenced on a SOLiD 5500×1 instrument according to the manufacturers' protocol. Variant calling was done using a custom pipeline as described in (Lips et al., submitted).

Sequence reads were mapped on the human reference genome version 19 (GRCh37), using BWA (Li and Durbin, 2009). To obtain copy number profiles from these targeted reads we used the CopywriteR tool (Kuilman et al. 2015, <https://github.com/PeeperLab/CopywriteR>). In brief, this tool extracts the off-target reads obtained with targeted

sequencing and uses these for copy number detection. The reads were then mapped to the BAC clone regions, and subsequently corrected for GC content and mappability and filtered for CNV regions as described above with the exception that mappability was corrected using a loess.

2.6. Mapping and BRCA1-like classification

The BRCA1-like classifier was originally trained on unsegmented BAC3K aCGH copy number profiles (Joosse et al., 2009; Vollebergh et al., 2011). The BRCA1-like classifier is a shrunken centroid classifier based on 371 (out of 3277) BAC clones (Vollebergh et al., 2011). For each platform we mapped raw copy number data-points to the 3277 BAC clones.

BAC3K, BAC32K, NG135, NG720, and SNP6 data were obtained as log2 copy number ratio; NGS, and targeted sequencing data were log2 read counts; MIP data was obtained as continuous copy number estimate (i.e. no ratio or log2). The MIP data was log2 transformed and subtracted by 1 to obtain 0-centered log2 values. We subsequently mapped these log2 ratio/value profiles to the original BAC3K aCGH platform on genome version hg18. Subsequently, we averaged the log2 ratios/values that fall within the chromosomal start and end position of the BAC clones (Schouten et al., 2013a). We used custom functions using the functionality from the following R packages in the mapping process: DNACopy (Venkatraman and Olshen, 2007), cghseg (Picard et al., 2011), Genomic Ranges (Aboyoun et al., 2013), KCSmart (De Ronde et al., 2009) and, copy number (Nilsen et al., 2012). The median BAC size was approximately 150 kb and the median number of probes averaged 3 for BAC32K (range 1–12), 6 for NG135 (range 1–30), 36 for NG720 (range 1–107), 92 for SNP6 (range 1–328), 15 for MIP (range 1–215) and the median number of 20 KB bins averaged for NGS was 8 (range 1–15). Missing BAC clones were filled by linear interpolation of the surrounding probes to obtain the 'mapped profile'. This mapped profile was classified to be BRCA1-like or non-BRCA1-like as described previously and used for all other further analyses (Schouten et al., 2013a; Vollebergh et al., 2011).

2.7. Statistical analysis

To evaluate the quality of each sample and to exclude low quality CN profiles we employed two statistical measures; 'variance of the noise' (VN) and 'signal to noise ratio' (SNR). The VN is defined as the variance between the processed signal (segmented copy number value) and the unprocessed signal (raw copy number value). The signal to noise is defined as the variance of the biological signal (log2 ratio of the segmented value) divided by the noise, as measured by the VN.

Profiles that had less signal than noise (SNR < 1) and high noise (VN > 0.025, as obtained from the density plot, Supplementary Figure 2) were considered low quality and excluded from the analysis. The similarity of samples analyzed by two platforms was visually assessed by plotting the average profile for each platforms. Hierarchical clustering was performed with a distance measure of 1-Pearson correlation and ward linkage. Subsequently, we checked whether repeated samples from the same patient clustered together.

For those samples we had BRCA1 methylation or mutation data available we calculated the sensitivity and the proportion of mutated/methylated samples in the BRCA1-like group. We calculated the inter-rater agreement between repeated samples using two measures and their respective confidence intervals: 1) the statistical accuracy values defined as the percentage of samples on the diagonal of the cross table of BRCA1-like status on one platform vs. the other platform and 2) Cohen's kappa value (R package *epiR* (Stevenson, 2012)). We used *Table1Heatmap* for plotting (Schouten, 2014). Cohen's kappa can be interpreted as follows: 0–0.4: poor agreement between tests; 0.4–0.8: moderate agreement between tests; and >0.8: near-perfect agreement (Schouten et al., 2013a). We calculated strength-of-classification and its standard deviation. Strength-of-classification is the Euclidean distance to the closest class. The value of this measure increases when a sample is closer and thus more strongly assigned to the class profile. 0 indicates that a sample is equally close to both classes. For example, if the Euclidean distance between a sample and the BRCA1-like average profile is 0.6, and the Euclidean distance between that sample and the non-BRCA1-like class is 0.9, the closest class is BRCA1-like and the measure is $0.9 - 0.6 = 0.3$. If for another sample the distance between a sample and the respective classes are 0.75 and 0.8, the measure would be 0.05, indicating less strong favor for any of the classes. Second, we used the standard deviation of the strength-of-classification. The larger the standard deviation the more likely a difference in classification.

All analyses were performed with R version 3.0.2.

3. Results

To establish the robustness of our BRCA1-like classifier on multiple CN platforms we used breast cancer samples that were analyzed by at least two genomic profiling platforms. The classifier, which was originally developed on a BAC aCGH platform, was tested on 263 tumor samples. The samples were analyzed using seven different technologies, with the overlap per technology ranging between zero (some platforms had no overlap) and 173 (NGS versus BAC3K) tumors. This resulted in 616 CN profiles, with 198 tumors overlapping between two, 43 between three, 19 between four, and two between five technologies. Forty profiles had an SNR smaller than 1 and a VN larger than 0.025 and were therefore removed, with another 31 profiles that lost a counterpart on another platform, resulting in 545 CN profiles spread over 230 patients. Table 1 describes the total number of profiles overlapping between two platforms and the number of profiles after quality control.

3.1. Mapped profiles resemble original profiles and biological signal overrules platform-specific characteristics

For every platform, we mapped the CN data to the BAC3K aCGH locations by averaging the log₂ ratios of positions overlapping each BAC clone, and investigated both the genome-wide and classifier region specific similarity between two platforms. We calculated the average genome-wide profile of samples that overlap per technology. Visual inspection

revealed high concordance between, segmented CN profiles (Figure 1), unsegmented CN profiles and CN profiles limited to the 371 classifier regions (Supplementary Figures 3 and 4). The distributions of the measurements were similar (Supplementary Figure 5). MIP, NGS and CopywriteR data demonstrated a larger dynamic range while Affymetrix SNP6 data displayed a compressed dynamic range. Within-sample variation (clustering by patient) is smaller than between-sample variation (clustering by technology) as is observed from hierarchical clustering analysis (Supplementary Figure 6,7). In conclusion, we observed high similarity between CN profiles after filtering low-quality genomic profiles and reducing dimension by averaging measurements that fall within a BAC clone. This is independent of the technology used.

3.2. BRCA1-like classification of mapped CN profiles is highly concordant with gold standard BAC3K classification

Having established the similarity between repeated samples from different datasets we investigated whether the minor differences in CN profiles would influence sample classification. We therefore performed a comparison between the tumor classification based on the original BAC3K profiles and profiles of the same tumors profiled on all other platforms. The class labels obtained from the BAC3K classifier served as gold standard. We then calculated the classification accuracy (how well does a classifier on another platform reproduce the BAC3K labels) and Cohen's kappa (what is the concordance between the BAC3K labels and the labels from another platform (Table 2).

The Cohen's kappa values between classification with mapped and original profiles ranged from 0.36 (BAC3K vs. CopywriteR) to 1 (BAC3K vs. BAC32K/NG720). With the exception of the MIP and CopywriteR results all Cohen's Kappa values are close to or above 0.8 indicating almost perfect concordance. However, some datasets are small resulting in wide confidence intervals (BAC3K vs. BAC32K), and some datasets do not have any samples classified differently, suggesting a potential bias of having only good quality and/or very concordant profiles in this particular analysis (BAC3K/NG720). A less stringent measure than Cohen's Kappa is the accuracy. For all technologies the percentage of samples that classify identically as the original profile was over 80%.

3.3. BRCA1-like classification is highly concordant with consensus classification across datasets

Although BAC3K is considered as the gold standard we should note that 1) the platform is not in operation anymore and 2) the gold standard classification may have been based on lower quality CN profiles (see below, section 3.7). Since we demonstrated high agreement for BRCA1-like classification of samples using the original BAC3K aCGH classification as a reference we investigated the overall agreement for each patient with all available data (Figure 2). In this analysis, we compared the class assigned with data from a particular platform to the class assigned based on the profiles from all other platforms.

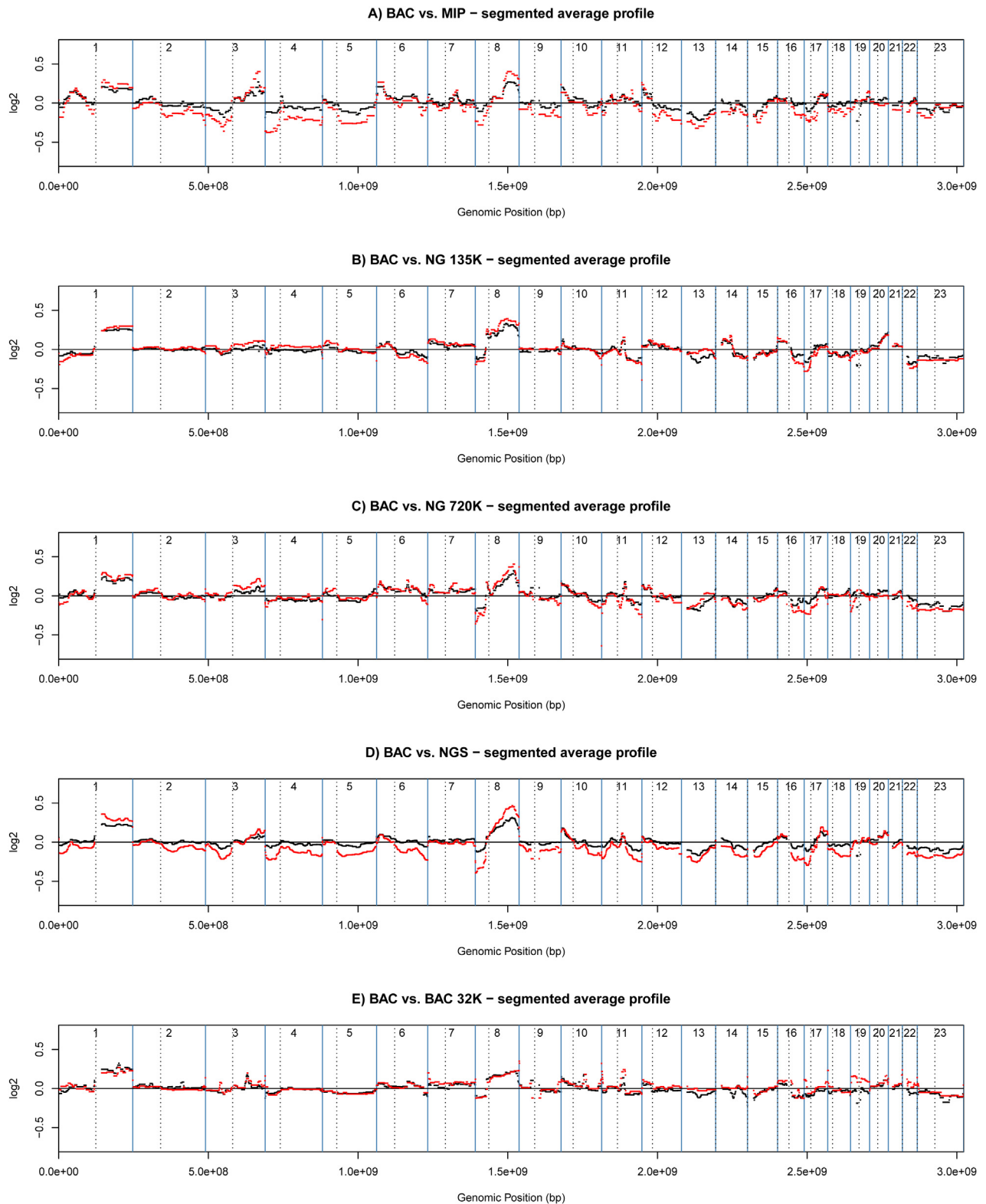


Figure 1 – Average copy number profiles compared between two technologies. Comparison of average of all samples based on their segmented copy number profiles showing the genomic position on the x-axis and the average log₂ ratio on the y-axis. Original profiles are plotted in black and mapped profiles in red. A) BAC vs. MIP segmented; B) BAC vs. NG135 segmented; C) BAC vs NG720 segmented; D) BAC vs. NGS segmented; E) BAC3K vs BAC32K segmented; F) NG135 vs. SNP6 segmented; G) CopywriteR vs BAC segmented; H) CopywriteR vs NG135 segmented.

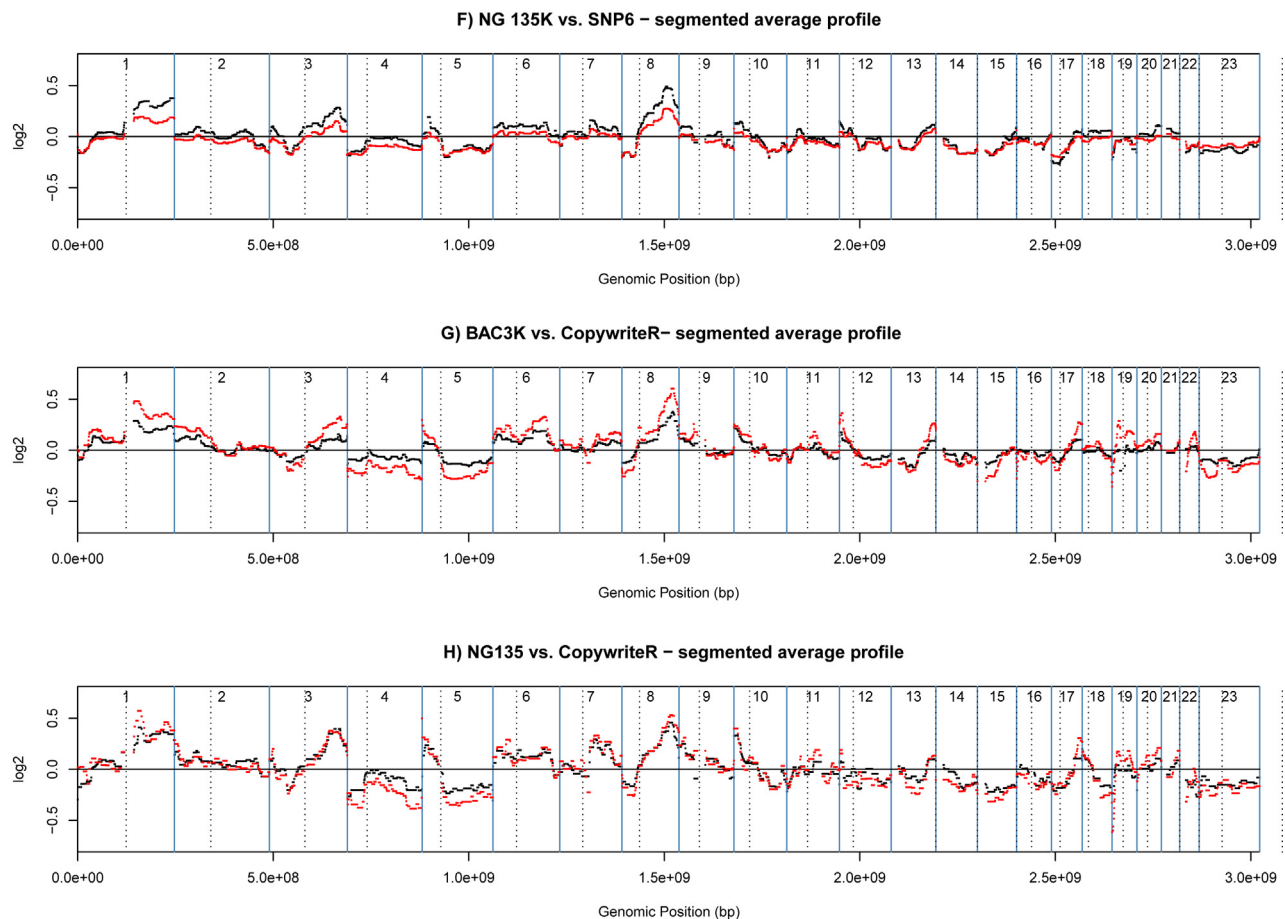


Figure 1 – (Continued)

206 out of 230 (90%, 95% CI of proportion: 85%–93%) samples had the same classification on all platforms, as far as they were profiled. As before, we calculated Cohen's kappa and accuracy values comparing one platform to all the others. In this analysis a sample was called BRCA1-like if it was BRCA1-like in any of the other platforms (Table 3, supplementary Table 1). Supplementary Figure 8 shows cross tables and kappa values for comparisons between all platforms. Accuracies of over 80% were obtained and kappa values of over 0.70. Only classification based on CopywriteR-extracted data had a lower kappa value (0.37), however, the accuracy remained high.

3.4. BRCA1-like status identifies BRCA1-mutated or –methylated cases

We investigated the performance in finding BRCA1-mutated and -methylated cases for the sequencing based datasets. In this series, the BAC classifier identified 89% (33/37), NGS 93% (28/30) and CopywriteR 100% (24/24) of the BRCA1-mutated or –methylated cases. BRCA1-like tumors were thus enriched for known causes of BRCA1 inactivation with respectively 33/48 (69%, BAC), 28/47 (60%, NGS) and 24/35 (69%, CopywriteR). The other cross tables are shown in Supplementary Table 2.

3.5. Sources of differential classification

Overall, classification between the tested platforms was similar. Subsequently, we performed a descriptive analysis to identify the causes underlying differential classification of samples (Supplementary Figure 9). We therefore re-analyzed including samples that were excluded due to quality control issues.

22 out of 31 patients that classified differently had CN profiles on two technologies, while eight had CN profiles on three and one on four technologies; 24 of these passed quality control with at least two profiles and were in the previous analysis. The shrunken centroid classifier will compare whether a sample is closer to the average profile of the BRCA1-like class or closer to the non-BRCA1-like class. We found that samples that have an inconsistent classification within one patient have a lower strength-of-classification and a larger standard deviation of strength-of-classification, compared to those patients that have the same classification (Figure 3). The filtering of samples with low quality partially removed this effect, indicating that samples with low signal and high noise are less strongly associated with a class. We observed that samples that failed to meet quality criteria were more likely those samples with different classification ($p = 0.04$, Fisher's exact test).

Table 2 – Cross tables of the classification of repeated samples with the proportion of samples classified the same (diagonal) and 95% confidence interval, Cohen's kappa and 95% confidence interval.

	BAC3K non-BRCA1-like	BAC3K BRCA1-like	accuracy	95%CI	Kappa	95%CI
BAC32K non-BRCA1-like	7	0				
BAC32K BRCA1-like	0	6	1.00	0.72–1	1.00	0.45–1.0
NG135K non-BRCA1-like	20	0				
NG135K BRCA1-like	1	2	0.96	0.76–1	0.78	0.37–1.0
NG720K non-BRCA1-like	11	0				
NG720K BRCA1-like	0	7	1.00	0.78–1	1.00	0.53–1.0
MIP non-BRCA1-like	1	0				
MIP BRCA1-like	2	8	0.81	0.47–0.97	0.42	0–0.90
NGS non-BRCA1-like	78	1				
NGS BRCA1-like	13	56	0.91	0.84–0.95	0.81	0.65–0.97
CopywriteR non-BRCA1-like	1	0				
CopywriteR BRCA1-like	3	21	0.88	0.68–0.97	0.36	0.06–0.66
	NG135K non-BRCA1-like	NG135K BRCA1-like				
SNP6 non-BRCA1-like	7	2				
SNP6 BRCA1-like	0	18	0.93	0.74–0.99	0.82	0.45–1.0
CopywriteR non-BRCA1-like	0	0				
CopywriteR BRCA1-like	2	10	0.83	0.51–0.97	N/A	N/A

Poor quality CN data could not explain all misclassifications, indicated by the fact that misclassification still occurred after quality control. Visual inspection of the profiles (Supplementary Figure 9) with different classifications demonstrated differences in quality between the profiles. Combining these findings we found a combination of three reasons for misclassification: 1) quality differences between CN profiles of the same patient, 2) lack of strong association in one of the CN profiles with one of the classes, 3) the presence of aberrations in genomic regions that are used for classification. The third point means that misclassification cannot occur if aberrations that are affected by differences in signal to noise ratio are absent from classifier regions. However, only a minority of samples is affected by a combination of these causes. Misclassification can be observed in Figure 2: weakly assigned cases have intermediate mean BRCA1-like probabilities and a small maximum difference between the BRCA1-like probabilities. Cases that are clearly discordant have a high maximum difference between the BRCA1-like probabilities. This could be due to either an incorrect original classification because of a lower BAC3K profile quality, or an incorrect classification based on the mapped profile. An incorrect mapped classification could occur because the original classifier was trained to recognize uncertainty in the classification (which results in probabilities around 0.5 for both classes) in the BAC3K profiles. Using mapped data with differences in, for example, noise or dynamic range could then increase the association with the wrong/correct class.

4. Discussion

In this study we investigated the robustness of our previously established BRCA1-like classification of CN profiles from breast cancer samples that were profiled on two to five different experimental platforms. We found that genomic position-based mapping between platforms results in

comparable CN profiles and subsequently similar BRCA1-like classification with high accuracy and agreement between platforms.

The overall comparability of CN profiling platforms for large aberrations (chromosome –arm level) (Baumbusch et al., 2008; Curtis et al., 2009; Krijgsman et al., 2012; Schouten et al., 2013a; Wicker et al., 2007) could be a beneficial characteristic for applying our BRCA1-like CN profile classifier (Joosse et al., 2009; Vollebergh et al., 2011). This classifier is based on genomic changes that arise in patients with BRCA1 mutation carriers or that have BRCA1 promoter methylation (Joosse et al., 2011, 2009; Lips et al., 2011; Vollebergh et al., 2011). We have demonstrated in three independent cohorts that this test could be used to predict benefit from high dose alkylating chemotherapy, which induces DNA double-strand breaks that cannot be repaired in BRCA1-deficient cancer cells (Schouten et al., 2015, 2014, 2013b; Vollebergh et al., 2011). Technological advances and the availability of many datasets have prompted an expansion of the methods deployed to obtain BRCA1-like classification of breast cancer samples.

Given that we did not aim to change the classifier itself, it is of importance to mimic the original BAC3K aCGH profiles as closely as possible. This limits the manipulations that can be done to improve the similarity between profiles. For example, we demonstrated previously (Schouten et al., 2013a) and in this manuscript that segmenting CN data results in repeated profiles that are more similar than unsegmented profiles. However, the segmentation negatively influences classification as the original classifier was trained on unsegmented data (Joosse et al., 2009; Schouten et al., 2013a; Vollebergh et al., 2011). As Curtis et al. described there is no ideal platform comparison because it is practically impossible to remove all differences between technologies when obtaining the actual copy number data (Curtis et al., 2009). In our case we also had experiments performed in the labs that were specialized in a certain technology. Therefore, it is not possible to derive the exact influence of the experimental platform and

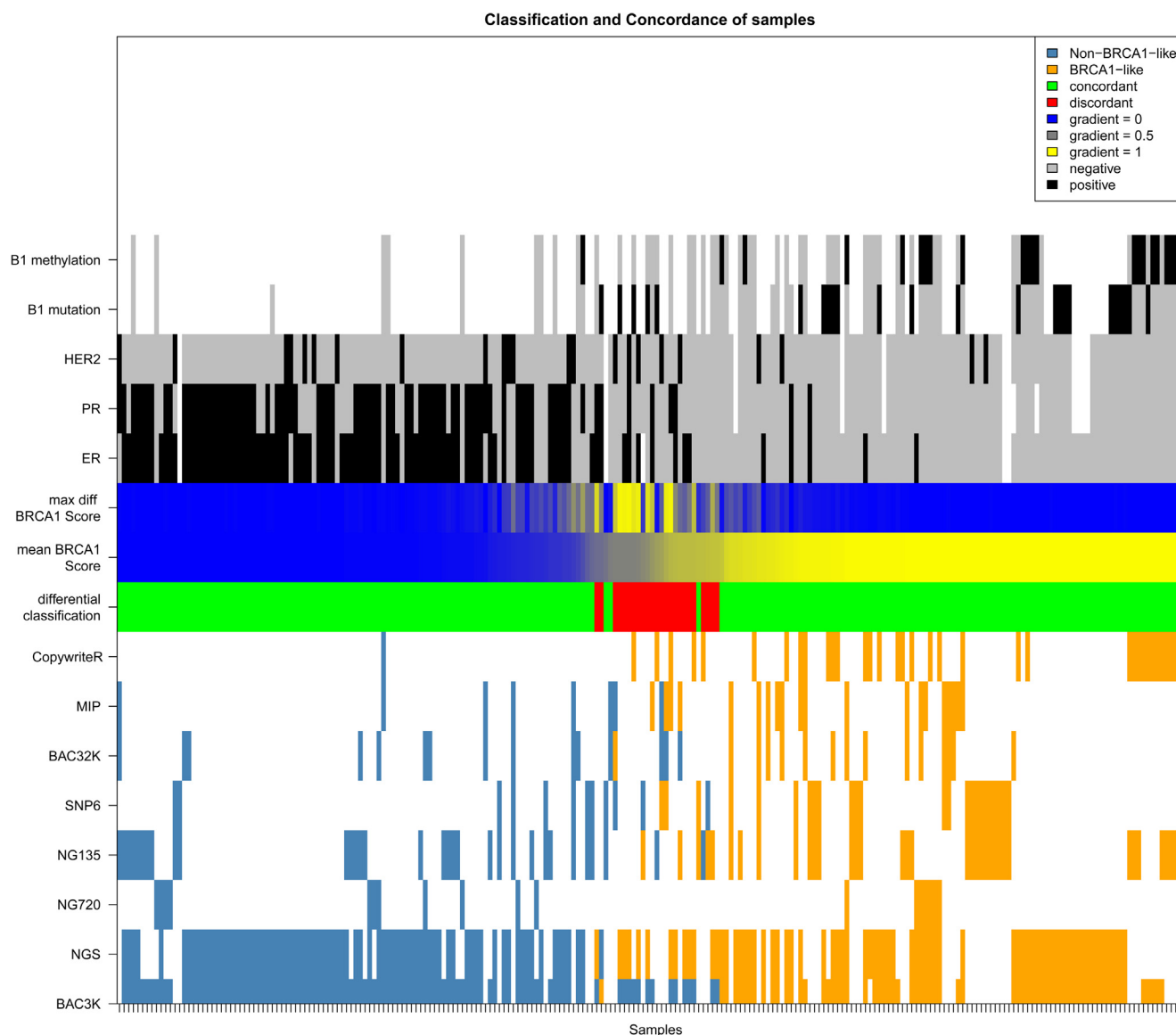


Figure 2 – Classification of all patients in the cohort. Classification of 210 patients with copy number data from 6 different copy number profiling platforms. Classification is colored as blue (non-*BRCA1*-like) and orange (*BRCA1*-like). The Differential classification row indicates whether differential classification occurred within a patient: green (no), red (yes). The mean *BRCA1*-like score is a gradient from blue (0) to yellow (1) indicating the probability of being *BRCA1*-like (score = 1). The max diff *BRCA1*-like score indicates the largest difference in probability between the classifications in one patient which ranges from 0 (no difference) to 1 (probability change of 1). These two scores can be used to identify clearly discordant samples (max diff → 1) or unconvincingly assigned profiles (mean score around 0.5, max diff small). ER, PR and HER2 status, *BRCA1* mutation status and *BRCA1* promoter hypermethylation status (negative or positive). Missing values are white.

experimental variations on classification. Furthermore, the applied measures of performance could be influenced by biological characteristics of the cohort. The biological characteristics of the cohort may confound estimating the performance. These characteristics alter the proportion of (non-)BRCA1-like tumors. For example, since BRCA1-like status associates with triple negative status, a cohort which contains mostly triple negative tumors (CopywriteR and MIP analysis) has few non-BRCA1-like cases. Few cases that misclassify result in a low kappa value while maintaining high accuracy (most concordantly classified to reference). An

equal amount of misclassifications in a cohort with more non-BRCA1-likes does not result in such a drop in kappa value. Using these two measures we hope to thus control for optimistic accuracy estimates with the kappa value, and to control for pessimistic estimates with accuracy, acknowledging that both can't control for confounding or a bias in selection of some of the cohorts. Keeping these limitations in obtaining copy number from different technologies and of the performance measures in mind, we observed a high concordance of classification when comparing to the original BAC aCGH-based classification. 7 out of 7 datasets reached the same

Table 3 – Cross tables, accuracy and concordance of classification for all pairs of samples for each profiling platform. *BRCA1*-like status was defined for each case if any of the repeated samples was *BRCA1*-like classified as *BRCA1*-like.

		Non- <i>BRCA1</i> -like	<i>BRCA1</i> -like	accuracy	95% CI	Kappa	95% CI
BAC3K	non- <i>BRCA1</i> -like	93	14	0.91	0.85–0.95	0.81	0.67–0.96
	<i>BRCA1</i> -like	1	58				
NGS	non- <i>BRCA1</i> -like	78	1	0.93	0.87–0.96	0.85	0.69–1.0
	<i>BRCA1</i> -like	11	58				
NG720	non- <i>BRCA1</i> -like	11	0	1	0.78–1	1	0.53–1
	<i>BRCA1</i> -like	0	7				
NG135	non- <i>BRCA1</i> -like	30	0	0.94	0.85–0.98	0.88	0.64–1
	<i>BRCA1</i> -like	2	25				
SNP6	non- <i>BRCA1</i> -like	9	3	0.89	0.73–0.96	0.74	0.41–1
	<i>BRCA1</i> -like	1	23				
BAC32K	non- <i>BRCA1</i> -like	12	3	0.87	0.68–0.95	0.73	0.38–1
	<i>BRCA1</i> -like	1	14				
MIP	non- <i>BRCA1</i> -like	6	2	0.89	0.71–0.97	0.73	0.35–1.00
	<i>BRCA1</i> -like	1	19				
CopywriteR	Non- <i>BRCA1</i> -like	1	0	0.92	0.76–0.98	0.37	0.12–0.63
	<i>BRCA1</i> -like	3	32				

classification in over 80% of the samples with MIP and CopywriteR possessing a low Cohen's kappa value for concordance. We attribute the lower kappa value for MIP and CopywriteR based data to the number and distribution of patients in the analysis as well as to differences in the quality of the profile or the mapping process. Mapping the MIP data to BAC clones requires transformation of CN to log2 ratio and a relatively high number of mapped MIP profiles did not meet the required quality criteria.

We tried to enlarge our datasets by pooling the data to obtain 'consensus classification' (in contrast to the gold standard) in a larger dataset. We found high accuracies (>0.8) in these pooled analyses, while kappa value was over 0.7 for

most comparisons. Only CopywriteR-based classification was low at 0.37. We attribute this to the distribution of patients (see above) and two cases that were not strongly assigned with probability scores around the cutoff.

In the direct comparison we used the *BRCA1*-like classification obtained by BAC aCGH as gold standard. Unfortunately, *BRCA1* mutation or methylation status was not available for all samples. For those datasets with available *BRCA1* status the classifier reproducibly found the *BRCA1*-mutated and methylated cases. This indicates that patients with a known *BRCA1*-inactivation are being identified. However, the classifier identifies *BRCA1*-like cases without *BRCA1*-mutation or –methylation that benefit from a chemotherapy regimen

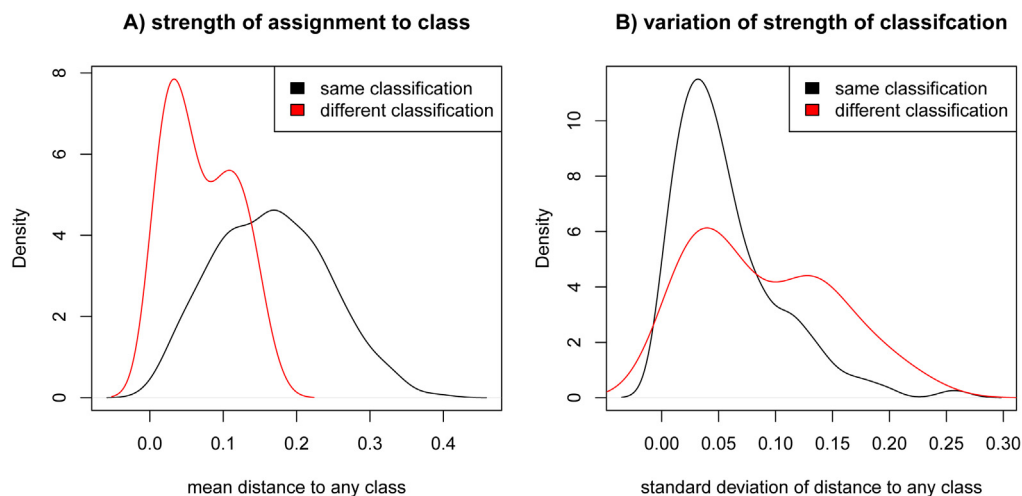


Figure 3 – Strength of classification Density plots of the strength and variation of classification within one patient for concordantly and discordantly classified patients. A) density plot indicating the strength of assignment to the *BRCA1*-like or non-*BRCA1*-like of all 243 patients. The strength of assignment is calculated as the mean absolute difference between one patient's profiles and the *BRCA1*-like and non-*BRCA1*-like class average profile. Zero indicates that the profile is equally close to the *BRCA1*-like as the non-*BRCA1*-like average. The higher the value the stronger its association with a particular class. B) Density plot of the standard deviation of the strength of assignment, indicating the association of the profiles from a patient with a particular class. In black are the patients for which all copy number profiles classified the same class, in red the patients that have different class assignment across technologies.

targeting the BRCA1 defect (Vollebergh et al., 2011; Schouten et al., 2015, 2014), which makes BRCA1-like classification the relevant read-out for predictive biomarker analyses.

We found that differential classification occurred in approximately 10% of the cases. The misclassification is caused by samples that lack strong characteristics of one class and/or varying quality profiles within a patient's samples. In cases with different classification it is uncertain which classification is correct, since the classifier may account for uncertainty arising from lower resolution BAC aCGH data. Conversely, it could be that lower quality BAC aCGH data obscured the true class and the use of higher resolution platform removed this uncertainty. In general, we observed high concordance and therefore applicability, at least for research purposes. The hypothetical ideal approach is always to optimize a classifier on the new platform with the same samples. However our results indicate that this may not be necessary. If optimizing is not feasible, similar classification can be obtained by mapping profiles. Furthermore, one should balance whether the benefit of optimizing the classifier on a platform outweighs (i.e. results in a much better classification) changing the test. It is advisable to obtain CN profiles for at least some tumor DNAs for which an original CN profile is available for comparison. Furthermore, a large reference set of profiles obtained with different platforms can be useful for finding outliers or potential errors in previous experiments (for example, our investigation of differential classification).

In conclusion, we demonstrated that BRCA1-like classification of mapped CN profiles is robust across multiple datasets and experimental platforms. This allows for the further investigation of the clinical benefit of treatments targeting the BRCA1 defect in existing datasets with CN profiles. Furthermore, the high concordance in CN profiles across different technologies encourages use on a range of current and established platforms.

Conflict of interest

SC Linn and PM Nederlof are named inventors on a patent application for the BRCA1 and -2-like array comparative genomics hybridization classifiers. The other authors do not disclose any conflict of interest.

Acknowledgments

This work was supported by an unrestricted research grant from Roche Life Science to translate the BRCA1-like classifier to a next generation sequencing platform, a grant from the Life Sciences Center Amsterdam (LSCA) Validation fund and the FP7 Collaborative Project "RATHER" (www.ratherproject.com).

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.molonc.2015.03.002>.

REFERENCES

- Aboyoun, P., Pages, H., Lawrence, M., 2013. GenomicRanges: Representation and Manipulation of Genomic Intervals. R Package Version 1.12.5.
- Alvarez, S., Diaz-Uriarte, R., Osorio, A., Barroso, A., Melchor, L., Paz, M.F., Honrado, E., Rodríguez, R., Urioste, M., Valle, L., Díez, O., Cigudosa, J.C., Dopazo, J., Esteller, M., Benitez, J., 2005. A predictor based on the somatic genomic changes of the BRCA1/BRCA2 breast cancer tumors identifies the non-BRCA1/BRCA2 tumors with BRCA1 promoter hypermethylation. *Clin. Cancer Res.* 11, 1146–1153.
- Baumbusch, L.O., Aarøe, J., Johansen, F.E., Hicks, J., Sun, H., Bruhn, L., Gunderson, K., Naume, B., Kristensen, V.N., Liestøl, K., Børresen-Dale, A.-L., Lingjaerde, O.C., 2008. Comparison of the Agilent, ROMA/NimbleGen and Illumina platforms for classification of copy number alterations in human breast tumors. *BMC Genomics* 9, 379.
- Bengtsson, H., Irizarry, R., Carvalho, B., Speed, T.P., 2008. Estimation and assessment of raw copy numbers at the single locus level. *Bioinforma* 24, 759–767.
- Benjamini, Y., Speed, T.P., 2012. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res* 40, e72. <http://dx.doi.org/10.1093/nar/gks001>.
- Buffart, T.E., Israeli, D., Tijssen, M., Vosse, S.J., Msić, A., Meijer, G.A., Ylstra, B., 2008. Across array comparative genomic hybridization: a strategy to reduce reference channel hybridizations. *Genes Chromosomes Cancer* 47, 994–1004.
- Cancer Genome Atlas Network, 2012. Comprehensive molecular portraits of human breast tumours. *Nature* 490, 61–70.
- Curtis, C., Lynch, A.G., Dunning, M.J., Spiteri, I., Marioni, J.C., Hadfield, J., Chin, S.-F., Brenton, J.D., Tavaré, S., Caldas, C., 2009. The pitfalls of platform comparison: DNA copy number array technologies assessed. *BMC Genomics* 10, 588.
- Curtis, C., Shah, S.P., Chin, S.-F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., Gräf, S., Ha, G., Haffari, G., Bashashati, A., Russell, R., McKinney, S., Langerød, A., Green, A., Provenzano, E., Wishart, G., Pinder, S., Watson, P., Markowitz, F., Murphy, L., Ellis, I., Purushotham, A., Børresen-Dale, A.-L., Brenton, J.D., Tavaré, S., Caldas, C., Aparicio, S., 2012. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 486, 346–352.
- De Ronde, J., Klijn, C., Velds, A., 2009. KCsmart: Multi Sample ACGH Analysis Package Using Kernel Convolution. <http://www.bioconductor.org/packages/release/bioc/html/KCsmart.html>.
- ENCODE Project Consortium, Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., Snyder, M., 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74.
- Esteller, M., Silva, J.M., Dominguez, G., Bonilla, F., Matias-Guiu, X., Lerma, E., Bussaglia, E., Prat, J., Harkes, I.C., Repasky, E.A., Gabrielson, E., Schutte, M., Baylin, S.B., Herman, J.G., 2000. Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. *J. Natl. Cancer Inst.* 92, 564–569.
- Harakalova, M., Mokry, M., Hrdlickova, B., Renkens, I., Duran, K., van Roekel, H., Lansu, N., van Roosmalen, M., de Bruijn, E., Nijman, I.J., Kloosterman, W.P., Cuppen, E., 2011. Multiplexed array-based and in-solution genomic enrichment for flexible and cost-effective targeted next-generation sequencing. *Nat. Protoc.* 6, 1870–1886. <http://dx.doi.org/10.1038/nprot.2011.396>.
- Hester, S.D., Reid, L., Nowak, N., Jones, W.D., Parker, J.S., Knudtson, K., Ward, W., Tiesman, J., Denslow, N.D., 2009. Comparison of comparative genomic hybridization technologies across microarray platforms. *J. Biomol. Tech.* 20, 135–151.

- Jooisse, S.A., Brandwijk, K.I.M., Devilee, P., Wesseling, J., Hogervorst, F.B.L., Verhoef, S., Nederlof, P.M., 2012. Prediction of BRCA2-association in hereditary breast carcinomas using array-CGH. *Breast Cancer Res. Treat* 132, 379–389.
- Jooisse, S.A., Brandwijk, K.I.M., Mulder, L., Wesseling, J., Hannemann, J., Nederlof, P.M., 2011. Genomic signature of BRCA1 deficiency in sporadic basal-like breast tumors. *Genes Chromosomes Cancer* 50, 71–81.
- Jooisse, S.A., van Beers, E.H., Nederlof, P.M., 2007. Automated array-CGH optimized for archival formalin-fixed, paraffin-embedded tumor material. *BMC Cancer* 7, 43.
- Jooisse, S.A., van Beers, E.H., Tielen, I.H.G., Horlings, H., Peterse, J.L., Hoogerbrugge, N., Ligtenberg, M.J., Wessels, L.F.A., Axwijk, P., Verhoef, S., Hogervorst, F.B.L., Nederlof, P.M., 2009. Prediction of BRCA1-association in hereditary non-BRCA1/2 breast carcinomas with array-CGH. *Breast Cancer Res. Treat* 116, 479–489.
- Kuilman, T., Velds, A., Kemper, K., Ranzani, M., Bombardelli, L., Hoogstraat, M., Nevedomskaya, E., Xu, G., de Ruiter, J., Lolkema, M.P., Ylstra, B., Jonkers, J., Rottenberg, S., Wessels, L.F., Adams, D.J., Peeper, D.S., Krijgsman, O., 2015. CopywriteR: DNA copy number detection from off-target sequence data. *Genome Biology* 16, 49. <http://dx.doi.org/10.1186/s13059-015-0617-1>.
- Krijgsman, O., Israeli, D., Haan, J.C., van Essen, H.F., Smeets, S.J., Eijk, P.P., Steenbergen, R.D.M., Kok, K., Tejpar, S., Meijer, G.A., Ylstra, B., 2012. CGH arrays compared for DNA isolated from formalin-fixed, paraffin-embedded material. *Genes Chromosomes Cancer* 51, 344–352.
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* 25, 1754–1760.
- Lips, E.H., Mulder, L., Hannemann, J., Laddach, N., Vrancken Peeters, M.T.F.D., van de Vijver, M.J., Wesseling, J., Nederlof, P.M., Rodenhuis, S., 2011. Indicators of homologous recombination deficiency in breast cancer and association with response to neoadjuvant chemotherapy. *Ann. Oncol.* 22, 870–876.
- Natrajan, R., Lambros, M.B., Rodríguez-Pinilla, S.M., Moreno-Bueno, G., Tan, D.S.P., Marchiò, C., Vatcheva, R., Rayter, S., Mahler-Araujo, B., Fulford, L.G., Hungermann, D., Mackay, A., Grigoriadis, A., Fenwick, K., Tamber, N., Hardisson, D., Tutt, A., Palacios, J., Lord, C.J., Buerger, H., Ashworth, A., Reis-Filho, J.S., 2009. Tiling path genomic profiling of grade 3 invasive ductal breast cancers. *Clin. Cancer Res.* 15, 2711–2722.
- Natrajan, R., Wilkerson, P.M., Marchiò, C., Piscuoglio, S., Ng, C.K.Y., Wai, P., Lambros, M.B., Samartzis, E.P., Dedes, K.J., Frankum, J., Bajrami, I., Kopeck, A., Mackay, A., A'hern, R., Fenwick, K., Kozarewa, I., Kumar, J., Mitsopoulos, C., Hardisson, D., Lord, C.J., Hakan-Sinha, C., Ashworth, A., Weigelt, B., Sapino, A., Chinnaiyan, A.M., Maher, C.A., Reis-Filho, J.S., 2014. Characterization of the genomic features and expressed fusion genes in micropapillary carcinomas of the breast. *J. Pathol.* 232, 553–565.
- Nilsen, G., Liestøl, K., Van Loo, P., Moen Volla, H.K., Eide, M.B., Rueda, O.M., Chin, S.-F., Russell, R., Baumbusch, L.O., Caldas, C., Børresen-Dale, A.-L., Lingjaerde, O.C., 2012. Copynumber: efficient algorithms for single- and multi-track copy number segmentation. *BMC Genomics* 13, 591.
- Picard, F., Lebarbier, E., Hoebeke, M., Rigai, G., Thiam, B., Robin, S., 2011. Joint segmentation, calling, and normalization of multiple CGH profiles. *Biostatistics* 12, 413–428.
- Rodenhuis, S., Bontenbal, M., Beex, L.V.A.M., Wagstaff, J., Richel, D.J., Nooij, M.A., Voest, E.E., Hupperets, P., van Tinteren, H., Peterse, H.L., TenVergert, E.M., de Vries, E.G.E., 2003. High-dose chemotherapy with hematopoietic stem-cell rescue for high-risk breast cancer. *N. Engl. J. Med.* 349, 7–16.
- Schouten, P.C., Marmé, F., Aulmann, S., Sinn, H.-P., van Essen, H.F., Ylstra, B., Hauptmann, M., Schneeweiss, A., Linn, S.C., 2015. Breast Cancers with a BRCA1-like DNA Copy Number Profile Recur Less Often Than Expected after High-Dose Alkylating Chemotherapy. *Clin. Cancer Res.* 21, 763–770. <http://dx.doi.org/10.1158/1078-0432.CCR-14-1894>.
- Schouten, P., 2014. Table1Heatmap. <http://cran.us.r-project.org/web/packages/Table1Heatmap/index.html>.
- Schouten, P.C., van Dyk, E., Braaf, L.M., Mulder, L., Lips, E.H., de Ronde, J.J., Holtman, L., Wesseling, J., Hauptmann, M., Wessels, L.F.A., Linn, S.C., Nederlof, P.M., 2013a. Platform comparisons for identification of breast cancers with a BRCA-like copy number profile. *Breast Cancer Res. Treat* 139, 317–327.
- Schouten, P., Gluz, O., Harbeck, N., Mohrmann, S., Diallo-Danebrock, R., Pelz, E., Kruizinga, J., Velds, A., Nieuwland, m, Kerkhoven, R., Liedtke, C., Frick, M., Kates, R., Linn, S., Marme, F., 2014. BRCA1-like copy number profiles to predict benefit of high-dose alkylating chemotherapy in high-risk breast cancer (BC): results from randomized WSG AM-01 trial. *J. Clin. Oncol.* 32 (5s) (suppl; abstr 11018).
- Schouten, P., Linn, S., Aulmann, S., Sinn, H., Schneeweiss, A., Marme, F., 2013b. BRCA1 like copy number profiles to predict benefit of intensified alkylating chemotherapy in breast cancer. *J. Clin. Oncol.* 2013 (suppl) abstr 11023.
- Stevenson, M., 2012. epiR: An R Package for the Analysis of Epidemiological Data.
- Tirkkonen, M., Johannsson, O., Agnarsson, B.A., Olsson, H., Ingvarsson, S., Karhu, R., Tanner, M., Isola, J., Barkardottir, R.B., Borg, A., Kallioniemi, O.P., 1997. Distinct somatic genetic changes associated with tumor progression in carriers of BRCA1 and BRCA2 germ-line mutations. *Cancer Res.* 57, 1222–1227.
- Tung, N., Miron, A., Schnitt, S.J., Gautam, S., Fettes, K., Kaplan, J., Yassin, Y., Buraimoh, A., Kim, J.-Y., Szász, A.M., Tian, R., Wang, Z.C., Collins, L.C., Brock, J., Krag, K., Legare, R.D., Sgroi, D., Ryan, P.D., Silver, D.P., Garber, J.E., Richardson, A.L., 2010. Prevalence and predictors of loss of wild type BRCA1 in estrogen receptor positive and negative BRCA1-associated breast cancers. *Breast Cancer Res.* 12, R95.
- Turner, N., Tutt, A., Ashworth, A., 2004. Hallmarks of “BRCAness” in sporadic cancers. *Nat. Rev. Cancer* 4, 814–819.
- Venkatraman, E.S., Olshen, A.B., 2007. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinforma* 23, 657–663.
- Venkitaraman, A.R., 2009. Linking the cellular functions of BRCA genes to cancer pathogenesis and treatment. *Annu. Rev. Pathol.* 4, 461–487.
- Vermaat, J.S., Nijman, I.J., Koudijs, M.J., Gerritse, F.L., Scherer, S.J., Mokry, M., Roessingh, W.M., Lansu, N., de Bruijn, E., van Hillegersberg, R., van Diest, P.J., Cuppen, E., Voest, E.E., 2012. Primary colorectal cancers and their subsequent hepatic metastases are genetically different: implications for selection of patients for targeted treatment. *Clin. Cancer Res.* 18, 688–699. <http://dx.doi.org/10.1158/1078-0432.CCR-11-1965>.
- Vollebergh, M.A., Jonkers, J., Linn, S.C., 2012. Genomic instability in breast and ovarian cancers: translation into clinical predictive biomarkers. *Cell. Mol. Life Sci.* 69, 223–245.
- Vollebergh, M.A., Klijn, C., Schouten, P.C., Wesseling, J., Israeli, D., Ylstra, B., Wessels, L.F.A., Jonkers, J., Linn, S.C., 2014. Lack of genomic heterogeneity at high-resolution aCGH between primary breast cancers and their paired lymph node metastases. *PLoS One* 9, e103177.
- Vollebergh, M.A., Lips, E.H., Nederlof, P.M., Wessels, L.F.A., Schmidt, M.K., van Beers, E.H., Cornelissen, S., Holtkamp, M., Froklage, F.E., de Vries, E.G.E., Schrama, J.G., Wesseling, J., van de Vijver, M.J., van Tinteren, H., de Bruin, M., Hauptmann, M., Rodenhuis, S., Linn, S.C., 2011. An aCGH classifier derived from BRCA1-mutated breast cancer and benefit of high-dose platinum-based chemotherapy in HER2-negative breast cancer patients. *Ann. Oncol.* 22, 1561–1570.

- Wang, Y., Moorhead, M., Karlin-Neumann, G., Falkowski, M., Chen, C., Siddiqui, F., Davis, R.W., Willis, T.D., Faham, M., 2005. Allele quantification using molecular inversion probes (MIP). *Nucleic Acids Res.* 33, e183.
- Wessels, L.F.A., van Welsem, T., Hart, A.A.M., van't Veer, L.J., Reinders, M.J.T., Nederlof, P.M., 2002. Molecular classification of breast carcinomas by comparative genomic hybridization: a specific somatic genetic profile for BRCA1 tumors. *Cancer Res.* 62, 7110–7117.
- Wicker, N., Carles, A., Mills, I.G., Wolf, M., Veerakumarasivam, A., Edgren, H., Boileau, F., Wasyluk, B., Schalken, J.A., Neal, D.E., Kallioniemi, O., Poch, O., 2007. A new look towards BAC-based array CGH through a comprehensive comparison with oligo-based array CGH. *BMC Genomics* 8, 84.